Structure of CYRI-B / FAM49B, a key regulator of cellular actin assembly

Elise Kaplan[a], Rachael Stone[a], Peter J. Hume[a], Nicholas P. Greene[a] and Vassilis Koronakis[a]*
[a]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, United Kingdom
Correspondence email: vk103@cam.ac.uk

Synopsis The crystal structure of CYRI-B is revealed, providing a template to understand the role of this highly conserved eukaryotic protein in a variety of actin dependent cellular processes.

Abstract In eukaryotes, numerous fundamental processes are controlled by the WAVE regulatory complex (WRC) that regulates cellular actin polymerization crucial to cell motility, cell-cell adhesion and epithelial differentiation. Actin assembly is triggered by interaction of the small GTPase Rac1 with CYFIP1, a key component of the WRC. Previously known as FAM49B, CYRI-B is a protein highly conserved across the Eukaryota which has recently been revealed to be a key regulator of Rac1 activity. Mutation, or alteration of CYRI-B expression, therefore, leads to altered actin nucleation dynamics with impacts on lamellipodia formation, cell migration and infection by intracellular pathogens. In addition, knockdown of CYRI-B expression in cancer cell lines results in accelerated cell proliferation and invasiveness. Here, the structure of *Rhincodon typus* (whale shark) CYRI-B is presented, the first to be reported of any CYRI family member. Solved by X-ray crystallography, the structure reveals that CYRI-B comprises three distinct α-helical subdomains and is highly structurally related to a conserved domain present in CYFIP proteins. The work presented here establishes a template towards a better understanding of CYRI-B biological function.

1. Introduction

Actin filament dynamics are central to a myriad of essential cellular processes such as cell migration, division and intracellular trafficking (Pollard & Cooper, 2009; Rottner *et al.*, 2017; Rottner & Schaks, 2019). Actin assembly is nucleated by cellular machines such as the ubiquitous Arp2/3 complex, which drives the generation of the branched actin networks underlying processes including lamellipodia formation and cell motility (Buracco *et al.*, 2019). Arp2/3 activity is regulated by nucleation-promoting factors (NPFs), which thus provide spatial and temporal control of these processes (Campellone & Welch, 2010; Rotty *et al.*, 2013). The best-characterised NPFs are those belonging to the Wiskott-

Aldrich syndrome protein (WASP) family, which stimulate Arp2/3 via C-terminal VCA (Verprolin-homology, Central and Acidic regions) domains (Pollitt & Insall, 2009; Alekhina *et al.*, 2017). The WASP family members WAVE1, WAVE2 and WAVE3 are central to cell motility and protrusion formation (Kurisu & Takenawa, 2009; Rottner & Schaks, 2019), and each functions as part of a heteropentameric complex termed the WAVE Regulatory Complex (WRC) (Chen *et al.*, 2010, 2014). In the resting state, the WAVE VCA domain is hidden in the WRC structure (Ismail *et al.*, 2009; Chen *et al.*, 2010). Multiple signals have been reported that can trigger exposure of the WAVE VCA and consequent activation of Arp2/3, the best-characterised of these being interaction with the small GTPase Rac1 (Miki *et al.*, 1998; Ismail *et al.*, 2009; Chen *et al.*, 2010), which is crucial for cell migration (Steffen *et al.*, 2013). There are two possible sites of Rac1 interaction on the WRC. The original 'A' ('adjacent') site is located in the N-terminal part of CYFIP1 (Chen *et al.*, 2010), a key component of the WRC, while the newly-identified 'D' ('distant') site is situated at the C-terminal extremity of the protein (Chen *et al.*, 2017). While two Rac1 molecules can simultaneously bind to the WRC *in vitro* (Chen *et al.*, 2017), only the 'A' site was found to be essential for its activation *in vivo* (Schaks *et al.*, 2018). Several studies reported that like CYFIP1, FAM49B interacts with the active form of Rac1 (Shang *et al.*, 2018; Fort *et al.*, 2018), a well-conserved eukaryotic protein. FAM49B was consequently renamed CYRI-B for CYFIP-related Rac1 interactor (Fort *et al.*, 2018). CYRI-B binding has been reported to block various Rac1-dependent signalling cascades in the cell, thereby controlling multiple critical cellular functions, including T cell activation (Shang *et al.*, 2018), membrane protrusion, chemotaxis and cell migration (Fort *et al.*, 2018; Whitelaw *et al.*, 2019). By negatively regulating Rac1 signalling, CYRI-B also reduces entry of several intracellular bacterial pathogens into both phagocytic and non-phagocytic host cells, and indeed is targeted for ubiquitin-mediated destruction by the action of an injected Salmonella virulence protein (Yuki *et al.*, 2019). CYRI-B has also been suggested to play a role in certain cancers. Thus, expression of CYRI-B in pancreatic ductal adenocarcinoma cells is downregulated in the tumour microenvironment, leading to increased cell proliferation and invasion (Chattaragada *et al.*, 2018). Furthermore, knockdown of CYRI-B expression increases cell proliferation in colorectal and liver cancer cell lines (Long *et al.*, 2019), suggesting CYRI-B may act as a tumour suppressor.

Here, we present the structure of *Rhincodon typus* (whale shark) CYRI-B, the first structure of any member of the CYRI protein family. Solved by X-ray crystallography at 2.4 Å resolution using the SAD method, the structure reveals three distinct, entirely α-helical, subdomains and a high structural similarity with a conserved domain of CYFIP1. Our work provides a basis to better understand the diverse and important roles of the CYRI protein family in eukaryotic cells.

2. Materials and methods

2.1. Cloning, expression and purification of CYRI-B (FAM49B)

DNA fragments encoding CYRI-B (FAM49B) from *Anthurium amnicola*, *Bison bison bison*, *Crocodylus porosus*, *Homo sapiens*, *Lingula anatine*, *Ornithorhynchus anatinus*, *Orussuss abientinus*, *Rhincodon typus*, *Tetranycus urticae* and *Tupaia chinensis* were synthesised (IDT) and cloned into pET11a using Gibson assembly (Gibson *et al.*, 2009). Fragments were designed to include an N-terminal Solubility Enhancement Tag (SET2, Stratagene), a hexahistidine tag (his-tag) and a Tobacco Etch Virus (TEV) cleavage site upstream of the CYRI-B coding sequence.

E. coli Rosetta (DE3) cells (Novagen) bearing the appropriate expression plasmid were grown at 37°C in 2YT media supplemented with 50 µg/mL of ampicillin and 20 µg/mL chloramphenicol. When the cells reached an OD_{600} of 0.8, the temperature was reduced to 18°C and protein expression was induced with 0.1 mM IPTG. After 16h at 18°C, bacteria were harvested by centrifugation at 6 000 g for 10 min at 4°C and the pellet resuspended in 40 mL of lysis buffer composed of 20 mM Tris pH 7.4, 150 mM NaCl, 0.25 mM Bond-Breaker TCEP solution (Thermo Scientific) and one tablet of EDTA-free protease inhibitor cocktail (Roche). Bacteria were lysed by cell disruption (Constant Systems) at 30 000 psi and the lysate clarified by ultracentrifugation for 1h at 60 000 g and 4°C. The clarified lysate was then incubated with 2 mL of Ni-NTA agarose (Qiagen) for 1h at 4°C. The resin was then washed with 20 mL of wash buffer (20 mM Tris pH 7.4, 500 mM NaCl, 0.25 mM TCEP) supplemented with 10 mM imidazole before elution with the same buffer containing only 150 mM NaCl and supplemented with 250 mM imidazole. The eluate was dialyzed against 5 L of dialysis buffer (20 mM Tris pH 7.4, 150 mM NaCl, 5 mM 2-mercaptoethanol). The N-terminal his-tag was removed using his-tagged Tobacco Etch Virus (TEV) protease produced from plasmid pTH24:TEVSH as described (Van Den Berg *et al.*, 2006). Cleavage reactions utilised 250 µM FAM49B and 185 µM TEV protease in a total volume of 10 mL and were incubated for 4h at 22ºC. Free his-tags, uncleaved CYRI-B protein and TEV protease were removed by several passages over 2 mL of Nickel resin (Biorad Profinity). Recombinant CYRI-B was concentrated to 10 mg/mL using a 10 kDa MWCO Amicon ultra centrifugal filter (Millipore), supplemented with 10% glycerol and stored at -80°C.

For preparation of selenomethionine (SeMet) derivatized CYRI-B, the metabolic inhibition protocol was used (Van Duyne *et al.*, 1993). *E. coli* Rosetta (DE3) cells carrying the CYRI-B expression plasmid were grown in M9 minimal medium supplemented with 50 µg/mL carbenicillin, 0.2% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 0.001% thiamine to an OD_{600} of 0.8. At this point the culture was supplemented with 60 mg/L selenomethionine, 50 mg/L leucine, isoleucine, valine and 100 mg/L lysine, phenylalanine and threonine. The temperature was reduced to 18ºC and cells were grown for a

further 15 minutes before induction of protein expression with 0.1 mM IPTG. After a further 16h, cells were pelleted by centrifugation at 6 000 g for 6 min. The SeMet labelled CYRI-B was then purified and the his-tag removed by TEV proteolysis as described for the native protein.

2.2. Analytical size-exclusion chromatography

Size-exclusion chromatography (SEC) was utilised to confirm the oligomeric state of the native and SeMet-derivatized CYRI-B proteins. SEC was performed at room temperature using an ÄKTA Pure FPLC system (GE Healthcare) equipped with a Superdex 75, 10/300 GL column in 20 mM Tris pH 7.5, 150 mM NaCl and 5 mM 2-mercaptoethanol. A total of 0.5 mg of native protein was applied on the column or 0.3 mg of the less soluble SeMet derivative, in a final volume of 100 µL.

2.3. Native and SDS-PAGE gels

The purity and oligomeric state of the recombinant CYRI-B proteins were analyzed using 7% native and 10% SDS-PAGE under non-reducing and reducing conditions respectively. Protein bands were visualised using Instant Blue™ (Expedeon).

2.4. Crystallization

Crystallization trials were set up in 2-drop MRC plates (Molecular Dimensions) using a Mosquito robot (TTP labtech). Crystals were grown using the sitting drop vapour diffusion method. Initial screening was performed using commercial crystallization screens (Structure Screen, MIDAS*plus*, JCSG-*plus* (Molecular Dimensions) and Salt RX (Hampton Research)). Subsequent optimisation utilised homemade screens.

Native CYRI-B protein at 10 mg/mL was mixed with the reservoir solution at a 1:2 ratio in a final volume of 1 µL over a reservoir of 80 µL. Crystals were obtained at 15°C in 0.2 M ammonium acetate, 0.1 M MES pH 6.5 and 30% (*w*/*v*) glycerol ethoxylate, using seeds of crystals obtained in 0.1 M MES pH 6.5, 7-10% PEG8000 and 20% PEG1000. Single orthorhombic prisms approximately 300 x 40 x 15 µm appeared the following day. Crystals were cryoprotected by the reservoir liquor and frozen in liquid nitrogen.

The SeMet-substituted CYRI-B was less soluble than the native protein and was consequently only concentrated to 2.6 mg/mL in a buffer containing 20 mM Tris pH 7.5, 150 mM NaCl and 5 mM 2-mercaptoethanol. Crystallization was carried out by mixing, at a 1:2 ratio, the protein with a precipitant solution composed of 0.2 M ammonium acetate, 0.1 M Bis-Tris, pH 5.5 and 25% PEG3350 in a final volume of 1 µL. Crystals were obtained after one day with the same seed stock used for the native protein but were slightly smaller (approximate size 50 x 20 x 10 µm). They were cryoprotected

with the reservoir solution supplemented with 10% (*v*/*v*) glycerol. Images of protein crystals together with a typical diffraction pattern are shown in Supplementary Figs. S1*a* and S1*b*.

2.5. Data collection and structure determination

The single-wavelength anomalous diffraction (SAD) data set was collected on beamline I03 at Diamond Light Source (DLS) from SeMet-labelled CYRI-B. A total of 720 images (0.04 s exposure, 0.5° oscillation) were collected on a Pilatus3 6M detector from a single crystal by illumination at 12 900 eV (0.9611 Å). The data were indexed and integrated with iMosflm (Battye *et al.*, 2011) in $P2_12_12_1$, scaled with Aimless (Evans & Murshudov, 2013) to a maximum resolution of 2.50 Å. The last 120 images were excluded due to radiation damage, as judged from batch I/σ(I) and R_{merge} statistics. Anomalous scatterers were located with the HySS (Hybrid Substructure Search, Grosse-Kunstleve & Adams, 2003) module of Phenix (Adams *et al.*, 2010; Liebschner *et al.*, 2019). The 15 heavy-atom sites were input into Phaser (McCoy *et al.*, 2007) for SAD phasing followed by density modification with Parrot (Cowtan, 2010) and model building with Buccaneer (Cowtan, 2006). The overall figure of merit for the final SAD phases was 0.48 rising to 0.57 after the first density modification. After each refinement step with Refmac (Murshudov *et al.*, 2011), model inspection and building was performed with Coot (Emsley *et al.*, 2010). The data were finally reprocessed at 2.40 Å and the model was refined with cycles of manual modification with Coot and refinement using Refmac first and Phenix for the last cycles including two TLS groups and automatic X-ray/stereochemistry weights. The final model has R_{work} and R_{free} factors of 0.2048 and 0.2415, respectively.

The native dataset was collected on beamline I04 at Diamond. The 720 images (0.05 s exposure, 0.5° oscillation) were collected on a Pilatus 6M-F detector. The data were processed with iMosflm in $P2_12_12_1$ and scaled with Aimless, giving a data set composed of 13 266 unique reflections to a maximum of 2.37 Å resolution. The structure was solved by molecular replacement with Phaser using the model obtained from the selenomethionine dataset. The model was then improved by several rounds manual building with Coot and refinement with Refmac first and Phenix for the last cycles using two TLS groups and automatic X-ray/stereochemistry constraints. The final model obtained has R_{work} and R_{free} factors of 0.2251 and 0.2772, respectively.

During refinement, eight ordered water molecules were manually added to the SeMet and native structures respectively. In both data sets, electron density was too weak to build residues 171 to 175. Both models were validated using tools from Coot, Rampage (Lovell *et al.*, 2003) and Procheck (Laskowski *et al.*, 1993). Data collection and structure refinement statistics are listed in Table 1. Coordinates and structure factors of the native and SeMet-derivatized proteins have been deposited in the Protein Data Bank (PDB) with the respective accession codes 6YJK and 6YJJ.

Table 1 Data collection and refinement statistics

Values for the outer shell are given in parentheses.

	SeMet CYRI-B 6YJJ	Native CYRI-B 6YJK
Data Collection		
Temperature (K)	100	100
Wavelength (Å)	0.9611	0.9795
Crystal Parameters		
Space Group	$P2_12_12_1$	$P2_12_12_1$
a, b, c (Å)	40.4, 72.8, 107.3	39.8, 72.2, 107.8
α, β, γ (°)	90, 90, 90	90, 90, 90
Reflection Data *		
Resolution (Å)	72.82-2.40 (2.49-2.40)	60.01-2.37 (2.46-2.37)
Unique Reflections	13006 (1329)	13266 (1362)
R_{sym}	0.112 (0.965)	0.122 (0.962)
I/σ(I)	12.8 (2.1)	10.1 (2.0)
$CC_{½}$	0.998 (0.782)	0.998 (0.590)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Multiplicity	9.5 (8.2)	10.4 (7.5)
Wilson B ($Å^2$)	44.2	47.1
Refinement †		
Resolution (Å)	53.66 (2.40)	60.01 (2.37)
Number of Reflections	12959	13207
R_{work} / R_{free}	0.2048 / 0.2415	0.2251 / 0.2772
Rms (Bond Lengths) (Å)	0.005	0.006
Rms (Bond Angles) (°)	0.70	0.84
Model Composition		
Protein atoms	2570	2573
Waters	8	8
Other	0	0
Model B-factors		
Protein atoms ($Å^2$)	63.2	74.5
Waters ($Å^2$)	42.5	45.5
Ramachandran Statistics ‡		
Favoured (%)	97.5	97.5
Allowed (%)	2.5	2.5
Outliers (%)	0.0	0.0

* Reflection data is as reported by Aimless (Evans & Murshudov, 2013).

† Refinement statistics as reported by Phenix (Adams *et al.*, 2010; Liebschner *et al.*, 2019).

‡ Ramachandran statistics as reported by Rampage (Lovell *et al.*, 2003).

2.6. Electrostatic potential calculations and figure preparation

Electrostatic potential profiles were calculated and visualized in PyMOL using the APBS Plugin (v2.1) (Baker *et al.*, 2001; Jurrus *et al.*, 2018) and PQR files generated by the PDB2PQR server (Dolinsky *et al.*, 2004, 2007). Surface potential was set to ± 4.0 kT/e. All figures were generated with PyMOL v2.2.2.

3. Results and discussion

3.1. Structure of CYRI-B (FAM49B), an α-helical protein

To facilitate structural studies, we cloned and expressed CYRI-B (FAM49B) from *Homo sapiens* and nine homologues. Most proteins either did not express well (*O. anatinus* and *T. chinensis*), were unstable after purification (*A. amnicola*, *B. b. bison* and *O. abientinus*) or failed, so far, to crystallize (*L. anatine*, *C. porosus*, *T. urticae* and *H. sapiens*). However, we successfully produced crystals of whale shark (*Rhincodon typus)* CYRI-B. After optimisation and screening of multiple crystals, we were able to obtain diffraction data at 2.37 Å resolution. We attempted to solve the structure by molecular replacement using the DUF1394 domain of CYFIP1 (PDB 3P8C) as it shares 21% sequence identity with whale shark CYRI-B and is predicted to be structurally similar (Yuki *et al*., 2019). However, these efforts were unsuccessful.

We therefore produced a selenomethionine-substituted protein that allowed us to determine the structure of CYRI-B by SAD phasing. The native and SeMet protein crystals have similar unit cell dimensions and belong to the same orthorhombic space group, $P2_12_12_1$. They were solved at resolutions of 2.37 and 2.40 Å, respectively. Collection and refinement data are listed in Table 1 and representative electron densities for the native and SeMet-substituted structures are respectively shown in Supplementary Movie S1 and Movie S2. In both cases, a single molecule is present in the asymmetric unit. The two models are nearly identical and could be aligned with a C_α r.m.s.d. of 0.29 Å for 320 atoms.

The structure of CYRI-B (FAM49B) is shown in Fig. 1. It reveals a protein comprised solely of α-helices which can be divided into three linear subdomains. The N-terminal subdomain extends until residue 125 and consists of five α-helices, three of which form a 60 Å elongated α-helical hairpin extending the entire length of the protein. The Medial subdomain extends from residues 126 to 214 and exhibits a 90º-broken antiparallel helical hairpin, that covers one face of the N-terminal subdomain. The two antiparallel helices α6 and α9 split at the lowest part to surround the extended α4 helix of the N-

terminal subdomain. Finally, the C-terminal subdomain (215-324) consists of six α-helices which associate into a globular bundle located at the base of the structure. This bundle caps the lower half of the Medial subdomain and is elevated above helices α1 and α5 of the N-terminal subdomain.

In both the native and SeMet structures, the flexible loop composed of residues 171 to 175 was too disordered to build and is therefore missing from the final models. This loop connects helices α7 and α8 of the Medial subdomain. A non-native serine, residual from the purification strategy, is present at the N-terminus of the structure.

CYRI-B possesses five cysteines, one is located in the N-terminal region while the C-terminal subdomain has four. In the C-terminal region, the two cysteines 231 and 253 from helices α10 and α12 are located too far apart (> 3 Å) to form an internal disulfide bond in our structure (Supplementary Fig. S2*a*). However, it is likely that in non-reducing conditions, these two helices are stabilised by a disulfide. Glycine 2 which is *N*-myristoylated *in vivo* (Fort *et al.*, 2018) is located at the base of the structure (Fig. 1) and is freely accessible.

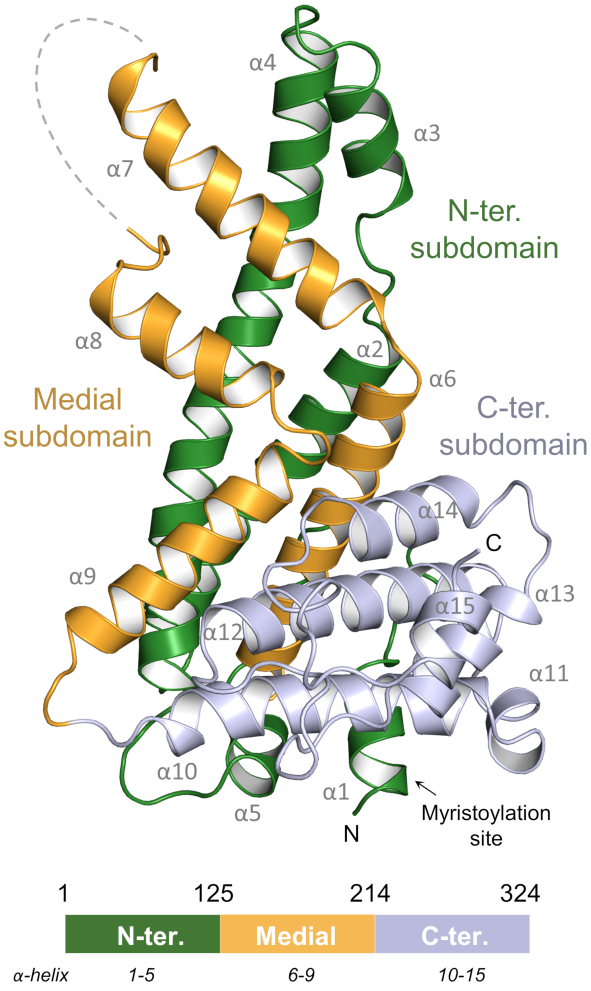


Figure 1 Crystal structure of CYRI-B. Cartoon representation of the native CYRI-B from *Rhincodon typus* showing the N-terminal (*green*), Medial (*yellow*) and C-terminal (*blue*) subdomains. The myristoylation site of the protein (Fort *et al.*, 2018) is indicated. Schematic linear organisation of the protein subdomains is represented underneath.

Biochemical analyses of *R. typus* CYRI-B were conducted to confirm the oligomeric state of the protein found by X-ray crystallography. As presented in Fig. 2*a*, the native and SeMet derivative proteins migrate as a single band on both native and SDS-PAGE gels. Under denaturing conditions, proteins migrate with an apparent molecular weight of ~ 30 kDa, close to the CYRI-B theoretical size of 37 kDa. Further analysis of the purified proteins by size-exclusion chromatography (SEC) revealed a single peak which elutes close to 30 kDa on a calibrated gel filtration column (Fig. 2*b*).

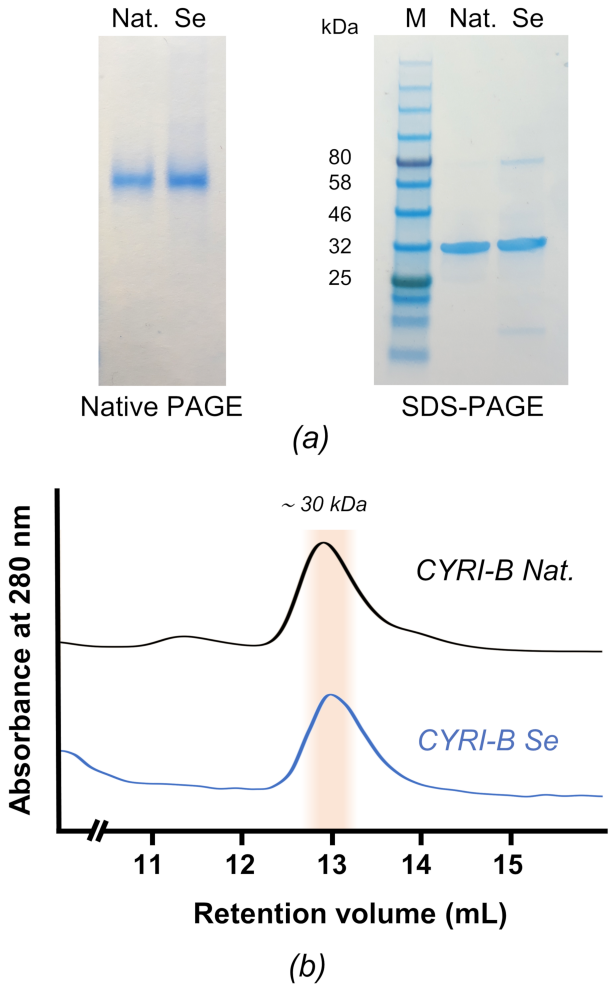


Figure 2 Oligomeric state of CYRI-B. (*a*) Native PAGE, *left*, and SDS-PAGE, *right*, gels of CYRI-B native (Nat.) and selenomethionine (Se) derivative proteins. Molecular masses of protein standards (M) are shown. (*b*) Size-exclusion chromatography profiles of indicated proteins.

Gel electrophoresis and SEC results corroborate the crystallographic data, confirming the monomeric state of CYRI-B in solution.

3.2. Comparison with human CYRI-B and relatives

3.2.1. Whale shark and human CYRI-B are almost indistinguishable

CYRI proteins are ubiquitous and highly conserved as illustrated by the alignment of sequences from multiple species (Supplementary Fig. S3*a*). For example, whale shark and human CYRI-B share a sequence identity of 93% (Supplementary Fig. S3*b*). Thus, over the 324 residues of both proteins, 12 substitutions are conservative, 5 semi-conservative and only 7 lead to a modification of the amino acid physico-chemical properties (Fig. 3*a*). Most of these substitutions localise to residues 10-18 which are located in a loop connecting helices α1 and α2 within the N-terminal subdomain (Fig. 3*b*). This region is poorly conserved in CYRI-B proteins (Supplementary Fig. S3*a*) and outside the presumed Rac1-binding region (Fort *et al.*, 2018). A few other mutations are located at the apex of this subdomain. The Medial and C-terminal subdomains are almost entirely conserved between the two species as they only contain two and three conservative substitutions respectively. As a result, the surface charge distribution of whale shark CYRI-B and the human homology model are near identical (Fig. 3*c* and Supplementary Movie S3).

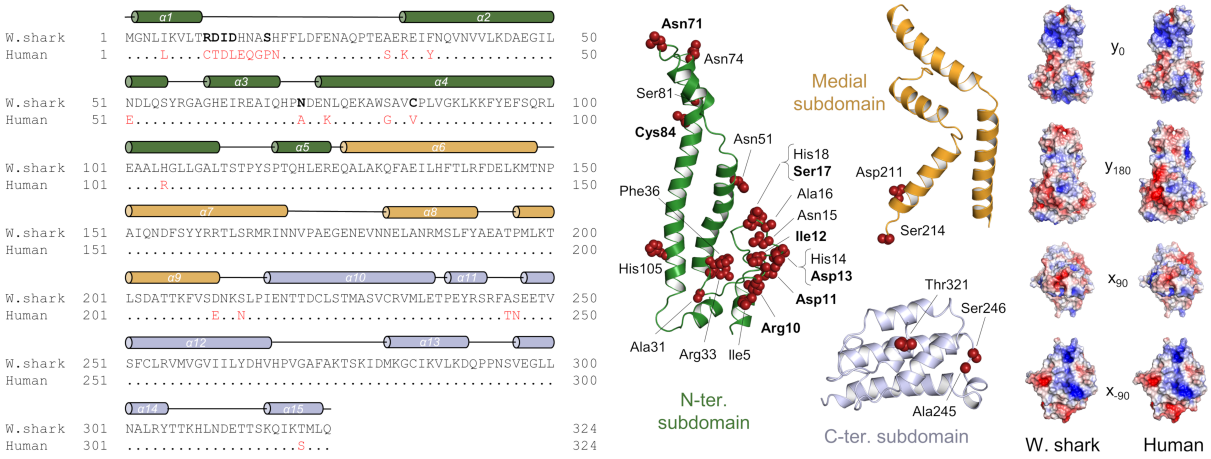


Figure 3 Comparison of human and whale shark CYRI-B reveals high sequence similarity. (*a*) Alignment of whale shark and human protein sequences where divergent amino acids are represented in *red*, and identical residues are symbolised by a dot. Secondary structure of whale shark CYRI-B is shown above the sequence, using the same colour code as in Fig. 1. Alignment was generated with the NCBI Blast program (Altschul *et al.*, 1997, 2005). (*b*) Side chains of dissimilar residues between the whale shark and human proteins are represented as spheres on each subdomain of the whale shark

CYRI-B structure. Non-conservative differences between the two proteins are shown in bold in (*a*) and (*b*). (*c*) Views of the electrostatic surface of whale shark, *left*, and human (Phyre homology model), *right*, CYRI-B proteins. A 360º rotation tour is presented in Supplementary Movie S3.

3.2.2. Comparison of CYRI-A and CYRI-B isoforms

Two isoforms of CYRI proteins exist in many species: CYRI-A and CYRI-B. Some cells express only CYRI-B, such as Jurkat T cells and others both (mouse T cells) (Shang *et al.*, 2018). The two proteins have a sequence identity of approximately 80% across multiple species including whale shark and human (Supplementary Fig. S3*b*), yet are functionally distinct. While a multitude of roles have been ascribed to CYRI-B, a clear function for CYRI-A has not been defined.

We compared the two isoforms of the whale shark proteins and mapped the thirteen substitutions which lead to modification of the amino acid physico-chemical properties on our CYRI-B structure (Fig. 4*a*). Most of them are located in the N-terminal subdomain, some on the flexible loop located between helices α1 and α2 and several along the extended α4 helix. Some of these substitutions modify the surface electrostatic potential (Fig. 4*b*). Thus CYRI-A is predicted to have a negatively-charged patch at the bottom of the structure where Gly109 in the -B isoform is replaced by a glutamate in the -A protein. Conversely, the presence of a glutamate (Glu73), absent in CYRI-A, leads to the formation of a negative patch at the apex of CYRI-B. Sequence analysis reveals these differences are conserved in CYRI proteins from other species (Fig. 4*c*) but until more is known about the function of CYRI-A, the significance of these differences remains unclear.

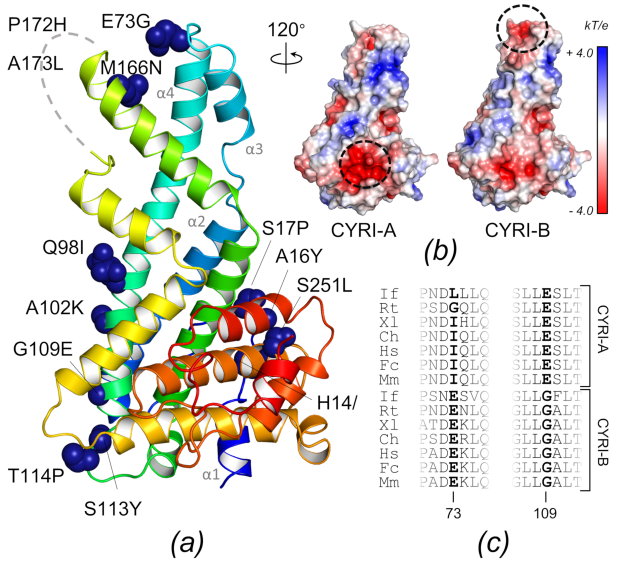


Figure 4 Comparison between whale shark CYRI-A and -B. (*a*) Radical substitutions between CYRI-A and -B as indicated by Clustal Omega (Sievers *et al.*, 2011) are shown as *blue* spheres on the CYRI-B structure. The corresponding residues in CYRI-B are indicated. (*b*) Electrostatic surface of CYRI-A

(Phyre homology model) and CYRI-B (crystal structure). (*c*) Sequence alignment of multiple CYRI-A and CYRI-B proteins highlighting residues 73 and 109 (*R. typus* CYRI-B numbering). Abbreviations are as follows: If, *I. furcatus;* Rt, *R. typus*; Xl, *X. laevis*; Ch, *C. horridus*; Hs, *H. sapiens*; Fc, *F. catus* and Ms, *M. musculus*. The full sequence alignment in provided in Supplementary Fig. S3*a*.

3.3. Similarity with CYFIP1 from the WAVE regulatory complex (WRC)

The ubiquitous CYFIP protein is a component of the WAVE regulatory complex (WRC). Two CYFIP isoforms exist, 1 and 2, which both contain a DUF1394 domain (residues 59-301, human numbering) also shared by CYRI proteins. Despite a relatively low sequence identity of 21% between human CYRI-B and the human CYFIP1 DUF1394 domain, the two regions are predicted to be structurally homologous (Yuki *et al.*, 2019). The comparison of CYFIP1 DUF1394 domain (PDB 3P8C) with our structure of CYRI-B confirms this prediction (Fig. 5*a*). The two proteins are strikingly similar and can be superposed with a r.m.s.d. score of 3.4 Å over 240 C_α and between 1.8 and 2.4 Å for the individual N-terminal, Medial and C-terminal subdomains (Fig. 5*b*).

Interestingly, the disordered loop (residues 169-181) connecting α7 and α8 is six residues longer than the corresponding region of CYFIP1 (residues 195-201). This loop constitutes one of the less conserved regions in CYRI-A and -B proteins (Supplementary Fig. S3*a*). Residues 171-175 which are missing in our structure could not be built based on the CYFIP1 structure as the flanking regions of the loop fold slightly differently in the two proteins (Supplementary Figs. S4*a* and S4*b*). Thus, helix α7 of CYRI-B is one turn longer than the corresponding helix in CYFIP1. The supplementary turn allows the interaction between Asn74 and the side chain of Gln318 belonging to a symmetry mate. Similarly, helix α8 of CYRI-B terminates one turn earlier, the loop unfolding towards the back of the structure where Asn177 interacts with the main chain of Asp290 belonging to a different symmetry mate. Moreover, the position of the loop in CYFIP1 is not compatible with the packing of CYRI-B crystal (Supplementary Fig. S4*c*). Indeed, the α4 apex of a symmetry mate would clash with an extra α-turn of α8 whereas the packing of the WRC crystal (PDB 3P8C) allows both conformations of this linker region.

We examined the conservation of spatially proximal cysteines 231 and 253 (see section 3.1), in CYRI proteins and CYFIP homologues. These cysteines are conserved in both CYRI-A and -B isoforms but in CYFIP homologues, only the equivalent of Cys231 is conserved (Supplementary Fig. S2*b*). Examination of CYFIP protein sequences and CYFIP1 structure (PDB 3P8C, Chen *et al.*, 2010) does

not reveal a plausible partner for Cys231. The significance of these conserved cysteines for the biological function of CYRI and CYFIP remains to be clarified.

The comparison of the two protein structures also highlights differences in the C-terminal subdomain. In CYRI-B, a loop connects α13 and α14 whereas in the equivalent region of CYFIP1, an 11 residue β-hairpin is present (Fig. 5*a*). This hairpin mediates both intra- and interdomain interactions in CYFIP1 (Fig. 5*c*, *bottom left*). Thus, the side chain of Asp310, located at the apex of the hairpin, forms a hydrogen bond with the Tyr84 hydroxyl of the N-terminal subdomain and main chain amides of Phe626 and Phe627 from the FragX-IP domain (residues 389-1222 of CYFIP1). Interaction with the FragX-IP domain is further stabilised by a hydrogen bond between Gln312 and the side chain of Ser413.

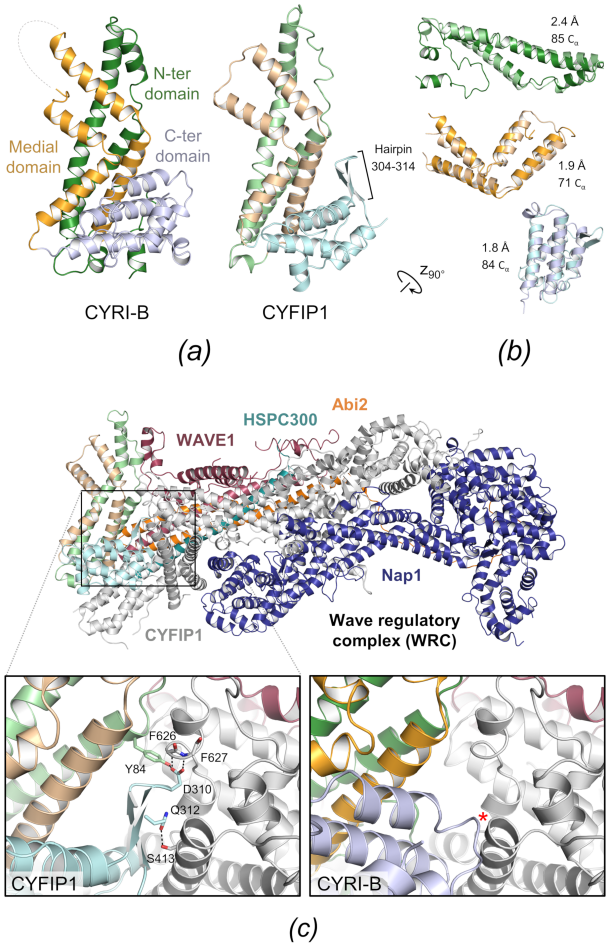


Figure 5 CYRI-B and CYFIP1 present a high degree of structural similarity. (*a*) Side-by-side view of the CYRI-B (*left*) and CYFIP1 DUF1394 domain (PDB 3P8C, *right*) structures. Colour code of CYRI-B is as Fig. 1. The presence of an extra hairpin in CYFIP1 is indicated. (*b*) Secondary structure alignment of the N-terminal (*green*), Medial (*yellow*) and C-terminal (*blue*) subdomains of the two proteins. The r.m.s.d. score and C_α number are shown beside each alignment (from Superpose, Krissinel

& Henrick, 2004). (*c*) *Top*, overall structure of the WAVE regulatory complex, WRC (PDB 3P8C). The DUF1394 domain of CYFIP1 is coloured as in (*a*). *Bottom left*, close-up view of the 304-314 hairpin of CYFIP1 showing residues involved in hydrogen interactions with the rest of the protein (*grey*). *Bottom right,* superposition of CYRI-B on the DUF1394 domain of CYFIP1 showing the absence of the 304-314 hairpin. An asterisk indicates a steric clash.

CYFIP1 has been shown to bind Rac1 (Chen *et al.*, 2010) and mutations that impair binding form a patch on CYFIP1 (Fig. 6*a*). Two of these mutations, C179R and R190D, are positioned on the DUF1394 domain of CYFIP1. Similarly, Shang and co-authors depicted the interaction of CYRI-B with Rac1 using co-immunoprecipitation and GST pull-down assays (Shang *et al.*, 2018). The association was reinforced with active forms of Rac1, *i.e.* the nucleotide-bound form and the constitutively activated G12V mutant. They subsequently identified a series of mutations which abolished the function of CYRI-B *in vivo*.

We mapped these mutations on our structure of CYRI-B (Figs. 6*a* and 6*b*). None of the residues are involved in intra-protein interactions, which is also the case for the two CYFIP1 mutants described to diminish the interaction between a modified version of the WRC and Rac1 (Fig. 6*b*) (Chen *et al.*, 2010).

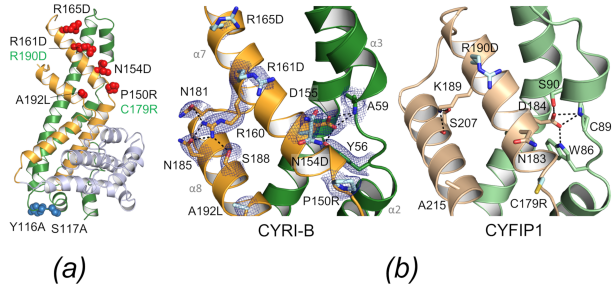


Figure 6 Predicted association of Rac1 with CYRI-B and CYFIP1. (*a*) Site directed mutagenesis data mapped onto the *R. typus* CYRI-B crystal structure. Mutations in CYRI-B (*black* label) that impair (*red* spheres) or do not affect (*blue* spheres) its function *in vivo* (Shang *et al.*, 2018) are shown. *Green* labels indicate mutations in the CYFIP1 DUF13243 domain that perturb the association of the WRC complex with Rac1 (Chen *et al.*, 2010). (*b*) Close-up views of the predicted Rac1-interacting regions of CYRI-B (*left*) and CYFIP1 (*right*), coloured as in Fig. 5. Residues leading to effective mutations in (*a*) are represented as *cyan* sticks with the corresponding mutated residue indicated. Electron densities of CYRI-B side chains are shown as *blue* mesh contoured at a sigma level of 1.

In both CYRI-B and CYFIP1, helix α7, which comprises the majority of the residues predicted to constitute the Rac1-binding region, is stabilised by interactions with the antiparallel helix α8 and

through a conserved aspartate that interacts with the loop joining helices α2 and α3 of the N-terminal subdomain (Fig. 6*b*, *left*). Thus, in CYRI-B, Asp155 forms hydrogen bonds with the hydroxyl of Tyr56 and the backbone amide of Ala59 while the equivalent CYFIP1 residue, Asp184, forms a H-bonding network with Trp86, Cys89 and Ser90 (Fig. 6*b*, *right*).

The highly conserved Arg161 of CYRI-B adopts an identical orientation to the equivalent residue in CYFIP1 (Arg190). This arginine is believed to play a central role in the association with Rac1 (Chen *et al.*, 2010; Fort *et al.*, 2018; Shang *et al.*, 2018). The side chain of a second conserved arginine (Arg160) in CYRI proteins forms extensive H-bonds with the side chain carbonyl groups of Asn181 and Asn185 and with the hydroxyl of Ser188, whereas the equivalent residue in CYFIP1 (Lys189) interacts with Ser207 carbonyl and hydroxyl groups. Interestingly, mutation of the conserved Arg160 to aspartate abolishes the interaction of CYRI-B with the activated Q61L Rac1 variant (Fort *et al.*, 2018). It is possible that this is due to a local disruption of the secondary structure as Arg160 stabilises the interaction of α7 and α8. Alternatively, the presence of two consecutive positively charged amino acids may be crucial for the recruitment of Rac1 in which case Arg160 may alter conformation to interact with specific residues of Rac1.

4. Conclusion

In summary, we have revealed the structure of CYRI-B, also called FAM49B, the only described negative regulator of cellular actin assembly mediated by the WAVE regulatory complex (WRC) (Fort *et al.*, 2018; Whitelaw *et al.*, 2019). Our structure, the first reported of any member of the CYRI family, reveals a protein entirely composed of α-helices and organised into three distinct subdomains. CYRI-B exhibits significant structural homology with CYFIP1, a component of the WRC that interacts with the small GTPase Rac1, to promote actin polymerisation (Bompard & Caron, 2004; Pollitt & Insall, 2009). The similarity in structures suggest CYRI-B interacts with Rac1 in the same manner as CYFIP1 and therefore regulates actin dependent processes by competition with the WRC for Rac1 binding. Residues previously described as vital for CYRI-B function (Shang *et al.*, 2018) cluster on a conserved helix which is likely to be central to Rac1 interaction. This study provides the structural basis to understand the function of CYRI proteins in a variety of fundamental cellular processes.

Acknowledgements The authors thank the staff at UK Diamond synchrotron for beamline access. We thank Susanna Berglund for the generous gift of plasmid pTH24:TEVSH and the MRC (MR/N000994/1) and Wellcome Trust (101828/Z/13/Z) for funding.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221.

Alekhina, O., Burstein, E. & Billadeau, D. D. (2017). *J. Cell Sci.* **130**, 2235–2241.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schaffer, A. A. & Yu, Y.-K. (2005). *FEBS J.* **272**, 5101–5109.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281.

Van Den Berg, S., Löfdahl, P. Å., Härd, T. & Berglund, H. (2006). *J. Biotechnol.* **121**, 291–298.

Bompard, G. & Caron, E. (2004). *J. Cell Biol. JCB J. Cell Biol.* **166**, 957–962.

Buracco, S., Claydon, S. & Insall, R. (2019). *F1000Research*. **8**, F1000 Faculty Rev:1977.

Campellone, K. G. & Welch, M. D. (2010). *Nat. Rev. Mol. Cell Biol.* **11**, 237–251.

Chattaragada, M. S., Riganti, C., Sassoe, M., Principe, M., Santamorena, M. M., Roux, C., Curcio, C., Evangelista, A., Allavena, P., Salvia, R., Rusev, B., Scarpa, A., Cappello, P. & Novelli, F. (2018). *Oncogene*. **37**, 697–709.

Chen, B., Chou, H. T., Brautigam, C. A., Xing, W., Yang, S., Henry, L., Doolittle, L. K., Walz, T. & Rosen, M. K. (2017). *Elife*. **6**, e29795.

Chen, B., Padrick, S. B., Henry, L. & Rosen, M. K. (2014). *Methods Enzymol.* **540**, 55–72.

Chen, Z., Borek, D., Padrick, S. B., Gomez, T. S., Metlagel, Z., Ismail, A. M., Umetani, J., Billadeau, D. D., Otwinowski, Z. & Rosen, M. K. (2010). *Nature*. **468**, 533–538.

Cowtan, K. (2006). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011.

Cowtan, K. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 470–478.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. (2007). *Nucleic Acids Res.* **35**, 522–525.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). *Nucleic Acids Res.* **32**, 665–667.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214.

Fort, L., Batista, J. M., Thomason, P. A., Spence, H. J., Whitelaw, J. A., Tweedy, L., Greaves, J., Martin, K. J., Anderson, K. I., Brown, P., Lilla, S., Neilson, M. P., Tafelmeyer, P., Zanivan, S., Ismail, S., Bryant, D. M., Tomkinson, N. C. O., Chamberlain, L. H., Mastick, G. S., Insall, R. H. & Machesky, L. M. (2018). *Nat. Cell Biol.* **20**, 1159–1171.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. (2009). *Nat. Methods*. **6**, 343–345.

Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 1966–1973.

Ismail, A. M., Padrick, S. B., Chen, B., Umetani, J. & Rosen, M. K. (2009). *Nat. Struct. Mol. Biol.* **16**, 561–563.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A. & Baker, N. A. (2018). *Protein Sci.* **27**, 112–128.

Krissinel, E. & Henrick, K. (2004). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268.

Kurisu, S. & Takenawa, T. (2009). *Genome Biol.* **10**, 226.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Crystallogr.* **26**, 283–291.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877.

Long, Y., Marian, T. A. & Wei, Z. (2019). *Biochem. Biophys. Res. Commun.* **513**, 1027–1034.

Lovell, S. C., Davis, I. W., Arendall, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins Struct. Funct. Genet.* **50**, 437–450.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658–674.

Miki, H., Suetsugu, S. & Takenawa, T. (1998). *EMBO J.* **17**, 6932–6941.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M.

D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367.

Pollard, T. D. & Cooper, J. A. (2009). *Science (80-.).* **326**, 1208–1212.

Pollitt, A. Y. & Insall, R. H. (2009). *J. Cell Sci.* **122**, 2575–2578.

Rottner, K., Faix, J., Bogdan, S., Linder, S. & Kerkhoff, E. (2017). *J. Cell Sci.* **130**, 3427–3435.

Rottner, K. & Schaks, M. (2019). *Curr. Opin. Cell Biol.* **56**, 53–63.

Rotty, J. D., Wu, C. & Bear, J. E. (2013). *Nat. Rev. Mol. Cell Biol.* **14**, 7–12.

Schaks, M., Singh, S. P., Kage, F., Thomason, P., Klünemann, T., Steffen, A., Blankenfeldt, W., Stradal, T. E., Insall, R. H. & Rottner, K. (2018). *Curr. Biol.* **28**, 3674–3684.

Shang, W., Jiang, Y., Boettcher, M., Ding, K., Mollenauer, M., Liu, Z., Wen, X., Liu, C., Hao, P., Zhao, S., McManus, M. T., Wei, L., Weiss, A. & Wang, H. (2018). *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4051–E4060.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.

Steffen, A., Ladwein, M., Dimchev, G. A., Hein, A., Schwenkmezger, L., Arens, S., Ladwein, K. I., Holleboom, J. M., Schur, F., Small, J. V., Schwarz, J., Gerhard, R., Faix, J., Stradal, T. E. B., Brakebusch, C. & Rottner, K. (2013). *J. Cell Sci.* **126**, 4572–4588.

Whitelaw, J. A., Lilla, S., Paul, N. R., Fort, L., Zanivan, S. & Machesky, L. M. (2019). *Commun. Integr. Biol.* **12**, 112–118.

Yuki, K. E., Marei, H., Fiskin, E., Eva, M. M., Gopal, A. A., Schwartzentruber, J. A., Majewski, J., Cellier, M., Mandl, J. N., Vidal, S. M., Malo, D. & Dikic, I. (2019). *Nat. Microbiol.* **4**, 1516–1531.

Supporting information

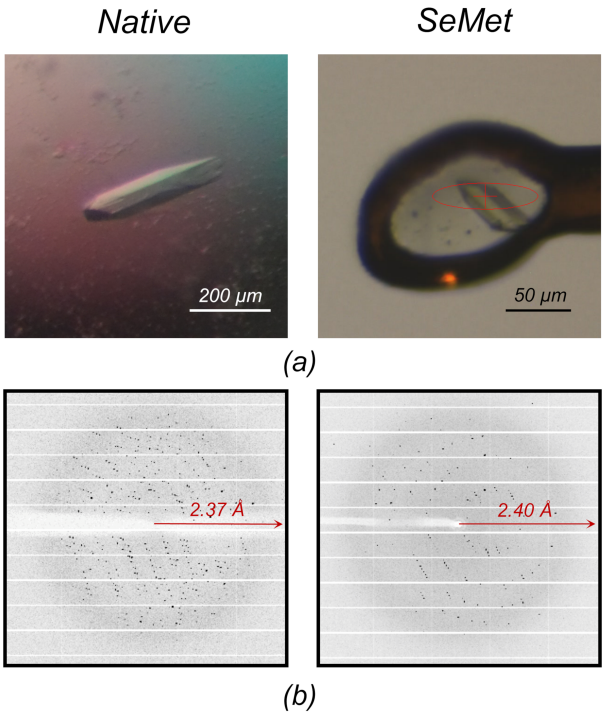


Figure S1 Crystallization and diffraction of *R. typus* CYRI-B (FAM49B). (*a*) Single orthorhombic prism-shaped crystals of native (*left*) and selenomethionine (SeMet, *right*) CYRI-B. (*b*) A typical diffraction pattern of native and SeMet CYRI-B (*left* and *right* respectively). The resolution limit is indicated.

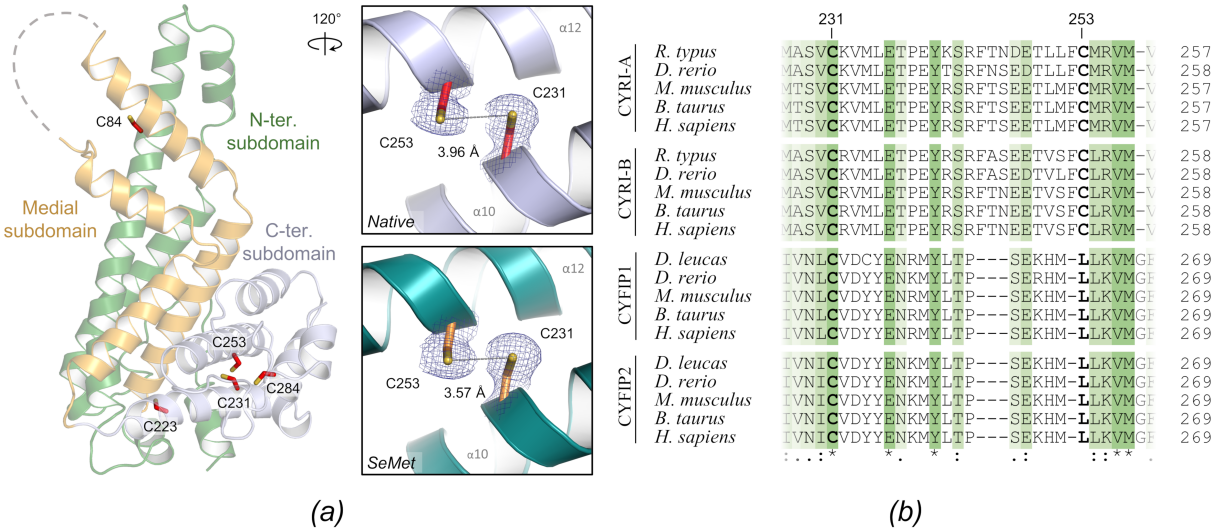
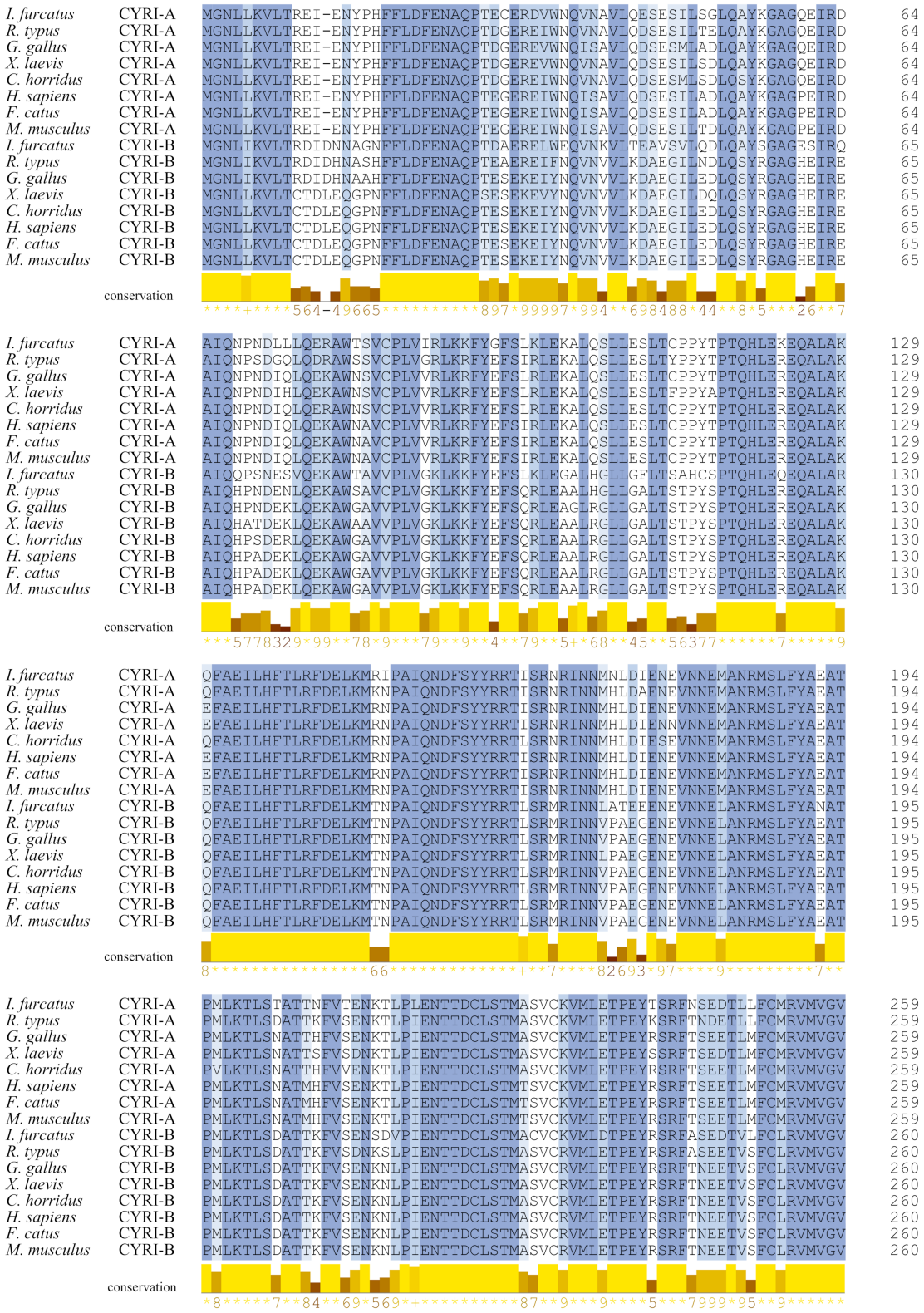


Figure S2 Location of cysteines in CYRI-B, showing a potential C231-C253 disulfide bond. (*a*, *left*) The five cysteines of CYRI-B are shown as *red* sticks on a cartoon representation of the *R. typus* crystal structure. (*a*, *right*) Close-up view showing electron density for the Cys231 and Cys253 side chains on the native (top) and selenomethionine (SeMet) derivative (bottom) proteins, crystallized in the presence of 2-mercaptoethanol. The mesh represents a weighted 2|Fo|-|Fc| electron density map contoured at 1 σ. The beta carbons of the two cysteines are 3.7 and 4.1 Å apart in the selenomethionine and native structure, respectively, compatible with the formation of a disulfide bond in non-reducing conditions. (*b*) Multiple sequence alignment of region 227 to 258 (*R. typus* CYRI-B numbering) of CYRI-A, CYRI-B, and 241 to 269 of CYFIP1 and CYFIP2 proteins. Alignment was performed with Clustal Omega (Sievers *et al.*, 2011) and the following sequences: *R. typus*, *Rhincodon typus* (NCBI XP_020388351.1 and XP_020368600.1); *D. rerio*, *Danio rerio* (Q5TZ57, Q6TLE5, Q90YM8 and A5A5E1); *M. musculus*, *Mus musculus* (Q8BHZ0, Q921M7, Q7TMB8 and Q5SQX6); *B. taurus*, *Bos taurus* (Q17QT7, Q2KJI3, E1BN47 and F1MX60); *H. sapiens*, *Homo sapiens* (Q9H0Q0, Q9NUQ9, Q7L576 and Q96F07); *D. leucas*, *Delphinapterus leucas* (A0A2Y9PLS0 and A0A2Y9PRF5). Except otherwise specified, sequences for CYRI-A, CYRI-B, CYFIP1 and CYFIP2 proteins respectively are from Uniprot. Annotation of *R. typus* genome did not allow clear identification of CYFIP proteins, which were substituted by CYFIP1 and CYFIP2 sequences of the related, Beluga whale (*D. leucas*). Residues are coloured according to the conservation symbols given by Clustal Omega.


I. furcatus CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTECERDVWNQVNAVLQESESILSGLQAYKGAGQEIRD 64
R. typus CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTDGEREIWNQVNAVLQDSESILTELQAYKGAGQEIRD 64
G. gallus CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTDGEREVWNQISAVLQDSESMLADLQAYKGAGQEIRD 64
X. laevis CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTDGEREVWNQVNAVLQDSESILSDLQAYKGAGQEIRD 64
C. horridus CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTDGEREIWNQVNAVLQDSESMLSDLQSYKGAGQEIRD 64
H. sapiens CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTEGEREIWNQISAVLQDSESILADLQAYKGAGPEIRD 64
F. catus CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTEGEREIWNQISAVLQDSESILADLQAYKGAGPEIRD 64
M. musculus CYRI-A MGNLLKVLTREI-ENYPHFFLDFENAQPTEGEREIWNQISAVLQDSESILTDLQAYKGAGPEIRD 64
I. furcatus CYRI-B MGNLIKVLTRDIDNNAGNFFLDFENAQPTDAERELWEQVNKVLTEAVSVLQDLQAYSGAGESIRQ 65
R. typus CYRI-B MGNLIKVLTRDIDHNASHFFLDFENAQPTEAEREIFNQVNVVLKDAEGILNDLQSYRGAGHEIRE 65
G. gallus CYRI-B MGNLIKVLTRDIDHNAAHFFLDFENAQPTESEKEIYNQVNVVLKDAEGILEDLQSYRGAGHEIRE 65
X. laevis CYRI-B MGNLLKVLTCTDLEQGPNFFLDFENAQPSESEKEVYNQVNVVLKDAEGILDQLQSYRGAGHEIRE 65
C. horridus CYRI-B MGNLLKVLTCTDLEQGPNFFLDFENAQPTESEKEIYNQVNVVLKDAEGILEDLQSYRGAGHEIRE 65
H. sapiens CYRI-B MGNLLKVLTCTDLEQGPNFFLDFENAQPTESEKEIYNQVNVVLKDAEGILEDLQSYRGAGHEIRE 65
F. catus CYRI-B MGNLLKVLTCTDLEQGPNFFLDFENAQPTESEKEIYNQVNVVLKDAEGILEDLQSYRGAGHEIRE 65
M. musculus CYRI-B MGNLLKVLTCTDLEQGPNFFLDFENAQPTESEKEIYNQVNVVLKDAEGILEDLQSYRGAGHEIRE 65
conservation
*****+****564-49665**********897*99997*994**698488*44**8*5***26**7
I. furcatus CYRI-A AIQNPNDLLLQERAWTSVCPLVIRLKKFYGFSLKLEKALQSLLESLTCPPYTPTQHLEKEQALAK 129
R. typus CYRI-A AIQNPSDGQLQDRAWSSVCPLVGRLKRFYEFSIRLEKALQSLLESLTYPPYTPTQHLEREQALAK 129
G. gallus CYRI-A AIQNPNDIQLQEKAWNSVCPLVVRLKRFYEFSIRLEKALQSLLESLTCPPYTPTQHLEREQALAK 129
X. laevis CYRI-A AIQNPNDIHLQEKAWNSVCPLVVRLKRFYEFSLRLEKALQSLLESLTFPPYAPTQHLEREQALAK 129
C. horridus CYRI-A AIQNPNDIQLQERAWNSVCPLVVRLKRFYEFSLRLEKALQSLLESLTCPPYTPTQHLEREQALAK 129
H. sapiens CYRI-A AIQNPNDIQLQEKAWNAVCPLVVRLKRFYEFSIRLEKALQSLLESLTCPPYTPTQHLEREQALAK 129
F. catus CYRI-A AIQNPNDIQLQEKAWNAVCPLVVRLKRFYEFSIRLEKALQSLLESLTCPPYTPTQHLEREQALAK 129
M. musculus CYRI-A AIQNPNDIQLQEKAWNAVCPLVVRLKRFYEFSIRLEKALQSLLESLTCPPYTPTQHLEREQALAK 129
I. furcatus CYRI-B AIQQPSNESVQEKAWTAVVPLVGKLKKFYEFSLKLEGALHGLLGFLTSAHCSPTQHLEQEQALAR 130
R. typus CYRI-B AIQHPNDENLQEKAWSAVCPLVGKLKKFYEFSQRLEAALHGLLGALTSTPYSPTQHLEREQALAK 130
G. gallus CYRI-B AIQHPNDEKLQEKAWGAVVPLVGKLKKFYEFSQRLEAGLRGLLGALTSTPYSPTQHLEREQALAK 130
X. laevis CYRI-B AIQHATDEKLQEKAWAAVVPLVGKLKKFYEFSQRLEAALRGLLGALTSTPYSPTQHLEREQALAK 130
C. horridus CYRI-B AIQHPSDERLQEKAWGAVVPLVGKLKKFYEFSQRLEAALRGLLGALTSTPYSPTQHLEREQALAK 130
H. sapiens CYRI-B AIQHPADEKLQEKAWGAVVPLVGKLKKFYEFSQRLEAALRGLLGALTSTPYSPTQHLEREQALAK 130
F. catus CYRI-B AIQHPADEKLQEKAWGAVVPLVGKLKKFYEFSQRLEAALRGLLGALTSTPYSPTQHLEREQALAK 130
M. musculus CYRI-B AIQHPADEKLQEKAWGAVVPLVGKLKKFYEFSQRLEAALRGLLGALTSTPYSPTQHLEREQALAK 130
conservation
5778329*99**78*979**9**4**79**5+*68**45**56377******7*****9
I. furcatus CYRI-A QFAEILHFTLRFDELKMRIPAIQNDFSYYRRTISRNRINNMNLDIENEVNNEMANRMSLFYAEAT 194
R. typus CYRI-A QFAEILHFTLRFDELKMKNPAIQNDFSYYRRTISRNRINNMHLDAENEVNNEMANRMSLFYAEAT 194
G. gallus CYRI-A EFAEILHFTLRFDELKMRNPAIQNDFSYYRRTISRNRINNMHLDIENEVNNEMANRMSLFYAEAT 194
X. laevis CYRI-A EFAEILHFTLRFDELKMRNPAIQNDFSYYRRTISRNRINNMHLDIENNVNNEMANRMSLFYAEAT 194
C. horridus CYRI-A QFAEILHFTLRFDELKMRNPAIQNDFSYYRRTLSRNRINNMHLDIESEVNNEMANRMSLFYAEAT 194
H. sapiens CYRI-A EFAEILHFTLRFDELKMRNPAIQNDFSYYRRTISRNRINNMHLDIENEVNNEMANRMSLFYAEAT 194
F. catus CYRI-A EFAEILHFTLRFDELKMRNPAIQNDFSYYRRTISRNRINNMHLDIENEVNNEMANRMSLFYAEAT 194
M. musculus CYRI-A EFAEILHFTLRFDELKMRNPAIQNDFSYYRRTISRNRINNMHLDIENEVNNEMANRMSLFYAEAT 194
I. furcatus CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNLATEEENEVNNELANRMSLFYANAT 195
R. typus CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNVPAEGENEVNNELANRMSLFYAEAT 195
G. gallus CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNVPAEGENEVNNELANRMSLFYAEAT 195
X. laevis CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNLPAEGENEVNNELANRMSLFYAEAT 195
C. horridus CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNVPAEGENEVNNELANRMSLFYAEAT 195
H. sapiens CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNVPAEGENEVNNELANRMSLFYAEAT 195
F. catus CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNVPAEGENEVNNELANRMSLFYAEAT 195
M. musculus CYRI-B QFAEILHFTLRFDELKMTNPAIQNDFSYYRRTLSRMRINNVPAEGENEVNNELANRMSLFYAEAT 195
conservation
8*****************66***********+**7****82693*97****9********7**
I. furcatus CYRI-A PMLKTLSTATTNFVTENKTLPLENTTDCLSTMASVCKVMLETPEYTSRFNSEDTLLFCMRVMVGV 259
R. typus CYRI-A PMLKTLSDATTKFVSENKTLPIENTTDCLSTMASVCKVMLETPEYKSRFTNDETLLFCMRVMVGV 259
G. gallus CYRI-A PMLKTLSNATTHFVSENKTLPIENTTDCLSTMASVCKVMLETPEYRSRFTSEETLMFCMRVMVGV 259
X. laevis CYRI-A PMLKTLSNATTSFVSDNKTLPIENTTDCLSTMASVCKVMLETPEYSSRFTSEDTLMFCMRVMVGV 259
C. horridus CYRI-A PVLKTLSNATTHFVVENKTLPIENTTDCLSTMASVCKVMLETPEYRSRFTSEETLMFCMRVMVGV 259
H. sapiens CYRI-A PMLKTLSNATMHFVSENKTLPIENTTDCLSTMTSVCKVMLETPEYRSRFTSEETLMFCMRVMVGV 259
F. catus CYRI-A PMLKTLSNATMHFVSENKTLPIENTTDCLSTMTSVCKVMLETPEYRSRFTSEETLMFCMRVMVGV 259
M. musculus CYRI-A PMLKTLSNATMHFVSENKTLPIENTTDCLSTMTSVCKVMLETPEYRSRFTSEETLMFCMRVMVGV 259
I. furcatus CYRI-B PMLKTLSDATTKFVSENSDVPIENTTDCLSTMACVCKVMLDTPEYRSRFASEDTVLFCLRVMVGV 260
R. typus CYRI-B PMLKTLSDATTKFVSDNKSLPIENTTDCLSTMASVCRVMLETPEYRSRFASEETVSFCLRVMVGV 260
G. gallus CYRI-B PMLKTLSDATTKFVSENKNLPIENTTDCLSTMASVCRVMLETPEYRSRFTNEETVSFCLRVMVGV 260
X. laevis CYRI-B PMLKTLSDATTKFVSENKNLPIENTTDCLSTMASVCRVMLETPEYRSRFTNEETVSFCLRVMVGV 260
C. horridus CYRI-B PMLKTLSDATTKFVSENKNLPIENTTDCLSTMASVCRVMLETPEYRSRFTNEETVSFCLRVMVGV 260
H. sapiens CYRI-B PMLKTLSDATTKFVSENKNLPIENTTDCLSTMASVCRVMLETPEYRSRFTNEETVSFCLRVMVGV 260
F. catus CYRI-B PMLKTLSDATTKFVSENKNLPIENTTDCLSTMASVCRVMLETPEYRSRFTNEETVSFCLRVMVGV 260
M. musculus CYRI-B PMLKTLSDATTKFVSENKNLPIENTTDCLSTMASVCRVMLETPEYRSRFTNEETVSFCLRVMVGV 260
conservation
*8*****7**84**69*569*+**********87**9***9****5***7999*95**9******

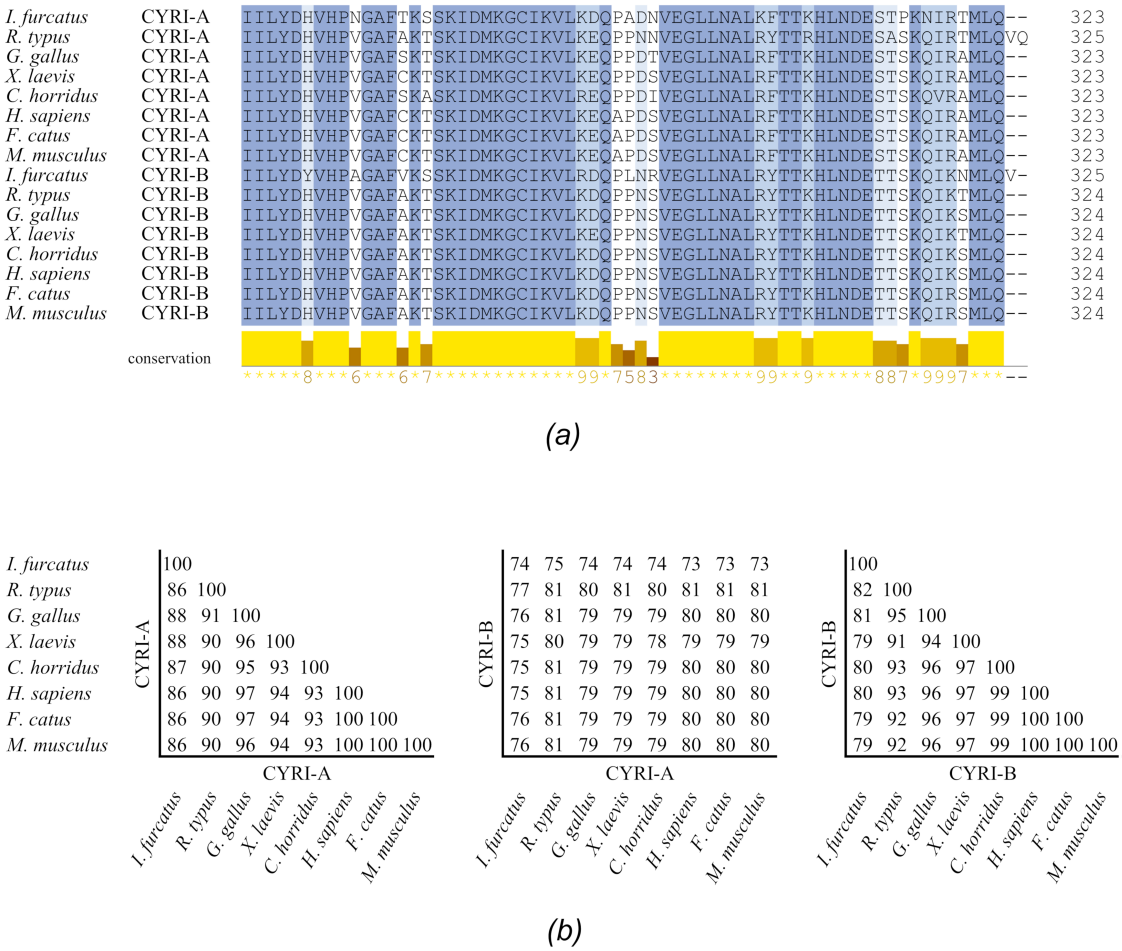


Figure S3 Sequence alignment of CYRI (FAM49) proteins across different species. (*a*) Alignment was generated with Clustal Omega (Sievers *et al.*, 2011) using the following protein sequences: *I. furcatus*, *Ictalurus furcatus* (Uniprot E3TCC8 and E3TCX7); *R. typus*, *Rhincodon typus* (NCBI XP_020388351.1 and XP_020368600.1); *G. gallus*, *Gallus gallus* (Uniprot Q5ZI04 and A0A3Q2TTC4); *X. laevis*, *Xenopus laevis* (Uniprot Q0IHF5 and Q8AVE9); *C. horridus*, *Crotalus horridus* (Uniprot A0A0K8RVD5 and A0A0B8RNP7); *H. sapiens*, *Homo sapiens* (Uniprot Q9H0Q0 and Q9NUQ9); *F. catus*, *Felis catus* (Uniprot A0A5F5XX33 and M3WMY8); *M. musculus*, *Mus musculus* (Uniprot Q8BHZ0 and Q921M7). Conservation at each (0, least conserved; 9 most conserved) was assessed in Jalview 1.0 and displayed beneath the sequence alignment. (*b*) Identity matrix based on the aligned sequences, generated by Clustal Omega.

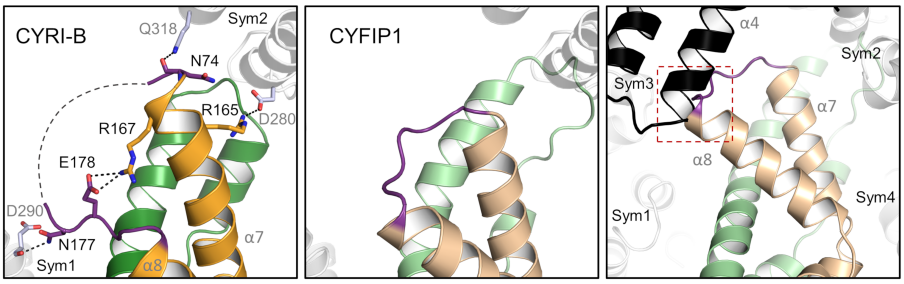


Figure S4 Close-up view of the disordered loop in *R. typus* CYRI-B. The loop that connects helices α7 and α8 is shown in *purple* in our crystal structure of CYRI-B (*a*) or in CYFIP1 (*b*, PDB 3P8C). It is composed of residues 169 to 181 in CYRI-B (195-201 in CYFIP1) and residues 171-175 were too disordered to model. Main local interactions between residues of the Medial subdomain and residues belonging to symmetry mate proteins are represented in CYRI-B structure. (*c*) Close-up view of CYFIP1 aligned on CYRI-B structure showing steric clash with one of the CYRI-B symmetry mates (*black*). A dotted *red* square highlights the position of the steric hindrance.

Movie S1 Representative electron density of the native CYRI-B (FAM49B) structure. A weighted 2|Fo|-|Fc| electron density map, calculated with model phases, is shown as *grey* mesh contoured at 1 σ.

Movie S2 Camera tour of the selenomethionine (SeMet) derivatized CYRI-B (FAM49B) structure. The first half of the movie shows the anomalous difference Fourier map surrounding selenium atoms as a *blue* mesh contoured at 3 σ. The second half of the film shows a weighted 2|Fo|-|Fc| electron density map, calculated with model phases, as *grey* mesh contoured at 1 σ.

Movie S3 Electrostatic surface of whale shark and human CYRI-B. Full 360º rotation movie around the y-axis showing the electrostatic surface of the whale shark CYRI-B crystal structure, *left*, and human homology model, *right*.